SUMMARIES OF ENTITIES (E.G., PEOPLE, PLACES, THINGS, CONCEPTS, ETC.) MAY PROVIDE ADDITIONAL USEFUL INFORMATION TO USER. FOR EXAMPLE, A SEARCH ENGINE MAY PROVIDE A SUMMARY OF AN ENTITY WITHIN SEARCH RESULTS. A CATEGORY (E.G., "WRITER", "POLITICIAN", ETC.) OF THE ENTITY THAT IS SHORT AND CONCISE MAY BE ADVANTAGEOUS TO PROVIDE WITHIN A SUMMARY OF THE ENTITY. THE CATEGORY MAY ALLOW A USER TO QUICKLY DETERMINE WHETHER THE INFORMATION OF THE ENTITY RELATES TO THE INTENDED ENTITY (E.G., SEARCH RESULTS OF AN ENTITY AS "A WRITER" VS. SEARCH RESULTS OF AN ENTITY AS "A POLITICIAN"). POTENTIAL CATEGORIES AND SUMMARY TEXT MAY BE EXTRACTED FROM PRE-LABELLED DATA. THE POTENTIAL CATEGORIES AND SUMMARY TEXT MAY BE INTERSECTED TO DETERMINE A SET OF CANDIDATE CATEGORIES THAT MAY BE RANKED. AN ENTITY CATEGORY HAVING A DESIRED RANKED MAY BE DETERMINED AS THE ENTITY CATEGORY THAT DESCRIBES THE ENTITY IN A DESIRED WAY.

```
100
START

102

104
EXTRACT ONE OR MORE POTENTIAL CATEGORIES AND SUMMARY TEXT RELATING TO ENTITY FROM PRE-LABELLED DATA

106
FOR RESPECTIVE CATEGORIES, INTERSECT CATEGORY WORDS OF POTENTIAL CATEGORY WITH SUMMARY WORDS WITHIN SUMMARY TEXT TO DETERMINE SET OF CANDIDATE CATEGORIES

108
RANK SET OF CANDIDATE CATEGORIES BASED UPON ONE OR MORE RANKING FEATURES

110
DETERMINE ENTITY CATEGORY FROM RANKED SET OF CANDIDATE CATEGORIES HAVING DESIRED RANK

END
```
100

START

102

EXTRACT ONE OR MORE POTENTIAL CATEGORIES AND SUMMARY TEXT RELATING TO ENTITY FROM PRELABELED DATA

104

FOR RESPECTIVE CATEGORIES, INTERSECT CATEGORY WORDS OF POTENTIAL CATEGORY WITH SUMMARY WORDS WITHIN SUMMARY TEXT TO DETERMINE SET OF CANDIDATE CATEGORIES

106

RANK SET OF CANDIDATE CATEGORIES BASED UPON ONE OR MORE RANKING FEATURES

108

DETERMINE ENTITY CATEGORY FROM RANKED SET OF CANDIDATE CATEGORIES HAVING DESIRED RANK

110

END

112

FIG. 1

AFTER HIS ESCAPE, MR. DAN LIVED IN HIDING UNTIL 1974.

SET OF PRE-LABELED CATEGORIES
- PRISON
- BRITISH SPIES FOR THE SOVIET UNION
- 1944 BIRTHS
- DOUBLE AGENT
- LIVING PEOPLE
- KGB OFFICERS

FIG. 3
CASH MANAGEMENT IS A NOVEL BY MRS. COLLEEN, FIRST PUBLISHED IN 1960 IN THE UNITED STATES. CASH MANAGEMENT IS AN ECONOMICS BOOK. MRS. COLLEEN IS FROM OHIO.

ENTITY: CASH MANAGEMENT

POTENTIAL CATEGORIES
- NOVEL BY MRS. COLLEEN
- ECONOMIC BOOKS
- RAIL TRANSPORT IN FICTION
- AMERICAN NOVELS
- FICTIONAL TECHNOLOGY
- WRITER AWARD WINNERS
- 1960 NOVELS
- CAPITALIST BOOKS
- BOOKS CRITICAL OF BIG BUSINESS
- OHIO BORN AUTHORS

FILTERING COMPONENT

SET OF CANDIDATE CATEGORIES
- NOVEL BY MRS. COLLEEN
- 1960 NOVELS
- ECONOMIC BOOKS
- OHIO BORN AUTHORS
- FICTIONAL TECHNOLOGY
- WRITER AWARD WINNERS

FIG. 4
SEARCH:
CASH MANAGEMENT

SEARCH RESULTS
SUMMARY FOR CASH MANAGEMENT
CATEGORY: NOVEL BY MRS. COLLEEN
- GENERA: ECONOMICS
- FIRST PUBLISHED: 1960

FIG. 6
FIG. 7

SEARCH ENGINE WEBSITE

SEARCH:

CASH MANAGEMENT NOVEL BY MRS. COLLEEN
FIG. 8
ENTITY CATEGORY DETERMINATION

BACKGROUND

[0001] The internet provides a rich source of information that users may search, consume, and/or share with others. For example, users may search for information regarding entities, such as people, places, things, blogs, companies, ideas, concepts, and/or a wide variety of other information. To enhance users' experience in searching and/or consuming content relating to entities, it may be advantageous to provide additional information, such as summaries, relating to the entities. In one example, a summary of an entity may comprise a category providing a short and succinct description of the entity (e.g., an entity of “Mr. Dan” may have a summary comprising a category of “writer”), which may aid a user in distinguishing “Mr. Dan” the “writer” from another Mr. Dan who is a politician). Given the vast amount of information available on the internet, however, deficiencies still remain regarding assisting users in finding relevant content.

SUMMARY

[0002] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key factors or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

[0003] Among other things, one or more systems and/or techniques for determining an entity category, at times making use of pre-labeled data, are disclosed herein. Entities may be interpreted as, among other things, people, places, things, blogs, companies, ideas, concepts, and/or a wide variety of ideas that may be found in an encyclopedia, for example. Pre-labeled data may be interpreted as information (e.g., web pages, articles, text, blogs, data stored in a database, and/or other electronic data) that may comprise pre-labeled categories associated with one or more entities. For example, pre-labeled data may be a web page comprising an encyclopedia type article about an entity, such as clouds, where that page also includes pre-labeled or pre-existing categories about that entity, such as weather, moisture, sun, rain, climate, etc.

[0004] Given that many categories may exist for a particular entity (e.g., an entity of “Mr. Dan” may have associated categories of both “writer” and “politician” (because Mr. Dan is both a writer and politician and/or there is a first Mr. Dan that is a writer and second Mr. Dan that is a politician)) and/or that many categories may be suitable or applicable to a particular entity, it may be desirable to weed out certain (potentially less relevant) categories, rank categories and/or otherwise select more suitable (potentially more relevant categories) for an entity. Accordingly, in an effort to achieve the same, as provided herein, one or more potential categories and summary text relating to an entity may be extracted from pre-labeled data. For example, pre-labeled data may comprise an article regarding clouds. Pre-labeled categories within the article may be extracted as potential categories relating to an entity of clouds. A portion of text within the article (e.g., a first occurring sentence, a first occurring paragraph, the entire article, etc.) may be extracted as summary text relating to the entity of clouds.

[0005] For respective potential categories, words of a potential category (e.g., category words) may be intersected with words within the summary text (e.g., summary words) to determine a set of candidate categories. That is, a potential category having a category word that appears in the summary text may be included in the set of candidate categories. In this way, potential categories may be filtered based upon the summary text. To take into account variations of words (e.g., singular vs. plural form), morphological analysis may be performed upon category words to generate variation category words that may be intersected with summary words within the summary text to determine the set of candidate categories. Additionally, syntactical features of the summary text may be determined based upon natural-language processing techniques (e.g., part-of-speech tagging, phrase chunking, etc.). One or more potential categories may be excluded from the set of candidate categories based upon the syntactical features.

[0006] The set of candidate categories may be ranked based upon one or more ranking features to generate a ranked set of candidate categories. For example, candidate categories may be ranked based upon a position within summary text feature (e.g., a position of one or more candidate words of a candidate category within the summary text in relation to verbs, prepositions, a first word within the summary text, etc.), display size feature (e.g., a text width of a candidate query, a font size, a bold format, etc.), and/or other candidate features. In one example, a candidate category may be assigned a desired rank based upon one or more candidate category words of the candidate category occurring after verbs (e.g., "is" or "was"). In another example, a candidate category may be assigned an undesired rank based upon one or more candidate category words of the candidate category occurring after prepositions (e.g., "from" or "in"). In this way, an entity category having a desired rank may be determined from the ranked set of candidate categories. In one example, the entity category may be presented as a search query completion suggestion. In another example, the entity category may be provided within a summary of an entity within a search result page.

DESCRIPTION OF THE DRAWINGS

[0007] To the accomplishment of the foregoing and related ends, the following description and annexed drawings set forth certain illustrative aspects and implementations. These are indicative of but a few of the various ways in which one or more aspects may be employed. Other aspects, advantages, and novel features of the disclosure will become apparent from the following detailed description when considered in conjunction with the annexed drawings.

[0008] FIG. 1 is a flow chart illustrating an exemplary method of determining an entity category.

[0009] FIG. 2 is a component block diagram illustrating an exemplary system for determining an entity category.

[0010] FIG. 3 is an illustration of an example of pre-labeled data.

[0011] FIG. 4 is an illustration of an example of determining a set of candidate categories for an entity.

[0012] FIG. 5 is an illustration of an example of ranking a set of candidate categories to generate a ranked set of candidate categories corresponding to an entity.

[0013] FIG. 6 is an illustration of an example of a search engine website presenting a summary comprising a category.

[0014] FIG. 7 is an illustration of an example of a search engine website providing a search query completion suggestion.
FIG. 8 is an illustration of an exemplary computer-readable medium wherein processor-executable instructions configured to embody one or more of the provisions set forth herein may be comprised.

FIG. 9 illustrates an exemplary computing environment wherein one or more of the provisions set forth herein may be implemented.

DETAILED DESCRIPTION

The claimed subject matter is now described with reference to the drawings, wherein like reference numerals are used to refer to like elements throughout. In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the claimed subject matter. It may be evident, however, that the claimed subject matter may be practiced without these specific details. In other instances, structures and devices are illustrated in block diagram form in order to facilitate describing the claimed subject matter.

When providing a summary of an entity, such as people, places, things, and/or concepts, it may be advantageous to provide a category description of the entity within the summary. The category may comprise a short and succinct description of the entity. For example, a category of an entity may have a category of “politician”. The category may help a user to understand what the entity is. For example, when a user begins to type a search query of an entity (e.g., “Marquette”), a category may be provided as a search query completion suggestion (e.g., “a city in Kansas”) to aid the user in narrowing the search to a desired entity (e.g., Marquette in Kansas, as opposed to Marquette in Michigan). Thus, it may be advantageous to leverage pre-labeled data having pre-labeled categories and summary text to determine an entity category that is short, concise, and highly relevant to the entity and to content that a user may be interested in consuming.

Accordingly, one or more systems and/or techniques for determining an entity category are provided herein. In particular, potential categories about an entity pre-labeled by manual input and summary text (e.g., a summary text, paragraph, and/or article summarizing an entity) may be extracted from pre-labeled data (e.g., a web page article regarding an entity). The potential categories and the summary text may be compared (e.g., intersected) to filter out undesirable potential categories. In this way, a set of candidate categories may be determined. The set of candidate categories may be ranked based upon ranking features (e.g., category word location within the summary text, category word size, etc.). An entity category having a desired rank may be determined from the ranked set of candidate categories (e.g., an entity category having a highest rank). The entity category may be used within a summary and/or within a search engine as a search query completion suggestion, for example.

One embodiment of determining an entity category is illustrated by an exemplary method 100 in FIG. 1. At 102, the method starts. At 104, one or more potential categories and summary text relating to an entity may be extracted from pre-labeled data. In one example, pre-labeled data may correspond to a web page article comprising a textual description of an entity (e.g., a portion of the textual description may be extracted as summary text, such as a first occurring sentence or a first occurring paragraph), and a set of pre-labeled categories relating to the entity. It may be appreciated that a potential category may comprise one or more category words (e.g., “Novel”, “by Mr. Dan”, “about”, “British spies”, “Novel by Mr. Dan” and/or other category words), where a category word may comprise a single word, number and/or symbol and/or a multiple words, numbers and/or symbols (e.g., a phrase of one or more words, one or more numbers and/or one or more symbols). It may be appreciated that category words may overlap or have overlapping terms (e.g., category word “Novel” and category word “Novel by Mr. Dan”). It may be appreciated that summary text may comprise one or more summary words (e.g., “Dan” “born in 1982” “raced sports cars”), where a summary word may comprise a single word, number and/or symbol and/or multiple words, numbers and/or symbols (e.g., a phrase of one or more words, one or more numbers and/or one or more symbols).

At 106, for respective potential categories, category words of a potential category may be intersected with summary words within the summary text to determine a set of candidate categories. For example, a potential category “raced sports cars” may be compared with summary text “Dan liked to race sports cars”. Because the category words “raced”, “sports”, “cars”, and/or variations thereof occur within the summary text, the potential category “raced sports cars” may be included within the set of candidate categories. In contrast, a potential category “born in 1982” may be compared with the summary text “Dan liked to race sports cars”. Because the category words “born in” and “1982” do not occur within the summary text, the potential category “born in 1982” may be excluded from the set of candidate categories.

Because a word may be used in multiple forms (e.g., singular vs. plural format), morphological or other analysis may be performed upon category words of a potential category to generate variation category words (e.g., race may be also analyzed instead of merely raced and vice versa). The variation category words of the potential category may be intersected with summary words within the summary text to determine the set of candidate categories (e.g., race may also be intersected with summary words instead of merely intersecting raced with summary words). Additionally, syntactical features of the summary text may be determined based upon natural-language processing techniques, such as part-of-speech tagging and/or phrase chunking. Potential categories may be excluded and/or included from/to the set of candidate categories based upon the syntactical features.

At 108, the set of candidate categories may be ranked based upon one or more ranking features. For example, ranking features may comprise a display size feature of category words within the summary text; a position within summary text feature corresponding to relative locations of category words in relation to other words (e.g., verbs, prepositions, etc.) within the summary text; etc. In one example, a desired rank (e.g., a high rank) may be assigned to candidate categories having category words occurring after verbs, such as “is” or “was”, within the summary text. In another example, an undesired rank (e.g., a low rank) may be assigned to candidate categories having category words occurring after prepositions, such as “from” or “in”, within the summary text. In this example, category words occurring after verbs may comprise more relevant and desired informa-
tion about the entity than category words occurring after prepositions. For example, when analyzing the text: Mr. Dan is a “carpenter” from “Ohio”, the fact that Mr. Dan is a “carpenter” may be more relevant and descriptive than the fact that Mr. Dan is from “Ohio”.

At an entity category having a desired rank may be determined from the ranked set of candidate categories. For example, a candidate category having a highest rank may be determined as the entity category. In one example, the entity category may be presented within a search result page in response to a user search query corresponding to the entity. In another example the entity category may be provided as a search query completion suggestion in response to user input comprising at least part of the entity’s name as a user search query. At the method ends.

In another example of determining an entity category corresponding to an entity, one or more potential categories may be extracted from an article of a web page. Summary text may be extracted from at least a portion of a first paragraph of the article. In one example, the web page, or a portion thereof, may be formatted or written in a certain style such that summary text can be extracted from the first paragraph, or a portion thereof, of the web page, for example (and the same may be true where summary text is obtained from additional/alternative portions the page). For respective potential categories, category words of a potential category may be intersected with summary words within the summary text to determine a set of candidate categories. Potential categories may be excluded from the set of candidate categories based upon syntactical features of the summary text. The set of candidate categories may be ranked based upon at least one of a position within summary text feature (e.g., a category word occurring after a preposition vs. a category word occurring after a verb) and a display size feature (e.g., font size of category words, length of category words, and/or other size features of category words within the summary text). An entity category having a desired rank may be determined from the ranked set of candidate categories. In this way, an entity category providing a short and concise summary of the entity may be determined.

FIG. 2 illustrates an example of a system configured to determine an entity category. The system may comprise an extraction component, a filtering component, a ranking component, and/or a presentation component. The extraction component may be configured to extract one or more potential categories and summary text relating to an entity from pre-labeled data. The filtering component may be configured to intersect category words of potential categories with summary words within the summary text to determine a set of candidate categories. The filtering component may be configured to perform morphological or other analysis upon category words of a potential category to generate variation category words. The filtering component may intersect the variation category words of the potential category with summary words within the summary text to determine the set of candidate categories. Additionally, the filtering component may be configured to determine syntactical features of the summary text based upon natural-language processing techniques, such as part-of-speech tagging and/or phrase chunking. The filtering component may be configured to exclude and/or include potential categories from/to the set of candidate categories based upon the syntactical features.

The ranking component may be configured to rank the set of candidate categories. In one example, the ranking component may assign a desired rank to a candidate category based upon one or more category words of the candidate category occurring after a verb, such as “is” or “was”. In another example, the ranking component may assign an undesired rank to a candidate category based upon one or more category words of the candidate category occurring after a preposition, such as “from” or “in”. In this example, category words occurring after verbs may comprise more relevant and desired information about the entity than category words occurring after prepositions. The ranking component may be configured to determine the entity category having a desired rank from the ranked set of candidate categories. The presentation component may be configured to present the entity category in response to user input comprising at least part of the entity’s name.

It will be appreciated that components of the system can be applied or enabled to varying degrees. For example, the ranking component and/or the filtering component may be throttled up or down to have a greater or lesser impact within the system. For example, if the ranking component 212 is throttled down to reduce (or eliminate) its effect, the entity category 214 may be determined mainly (or completely) based upon the implementation of the filtering component 208. Similarly, if the filtering component 208 is turned off to reduce (or eliminate) its effect, the entity category 214 may be determined mainly (or completely) based upon the implementation of the ranking component 212.

FIG. 3 illustrates an example of pre-labeled data. The pre-labeled data may correspond to a website that provides articles on various topics, such as entities (e.g., people, places, things, concepts, etc.). For example, the pre-labeled data may correspond to an article on Mr. Dan (an entity of Mr. Dan). The article may comprise text describing Mr. Dan, an image of Mr. Dan, a set of pre-labeled categories, and/or other information relating to Mr. Dan. A portion of the text may be extracted as summary text. Additionally, the set of pre-labeled categories may be extracted as one or more potential categories.

It may be appreciated that a filtering component (e.g., filter component 208 of FIG. 2) may intersect category words of potential categories with summary words within the summary text to determine a set of candidate categories. For example, “living people”, “KBG officers”, and/or other potential categories may be excluded from a set of candidate categories because category words of the potential categories do not match summary words within the summary text. In contrast, because category word “British spies” and category word “Soviet spies” do not match summary words within the summary text, “British spies” and “Soviet spies” may be included in the set of candidate categories because at least some of the category words match summary words within the summary text. For example, “British spies” may match “British Spy” and “Soviet Union” may match “Soviet Union”. Other potential catego-
ries, such as “prison”, “1944 births”, and/or “double agent” may likewise be included within the set of candidate categories, for example.

In one example, the ranking component 516 may rank a candidate category “novel by Mrs. Colleen” with a desired rank (e.g., a high rank) because candidate words of the candidate category “novel by Mrs. Colleen” occur after a verb “is” within summary text 502 (e.g., “is a novel by Mrs. Colleen” 504). The ranking component 516 may rank a candidate category “1960 novels” with an undesired rank (e.g., a medium/low rank) because candidate words of the candidate category “1960 novels” occur after a preposition “in” within the summary text 502 (e.g., “in 1960” 506).

The ranking component 516 may rank a candidate category “economic books” with a desired rank (e.g., a very high rank) because candidate words of the candidate category “economic books” occur after a verb “is” within summary text 502 and the candidate words have a large display size/font within the summary text 502 (e.g., “is an economics book” 508). In this way, the ranking component 516 may rank the set of candidate categories 518 to generate the ranked set of candidate categories 520. For example, the ranked set of candidate categories 520 may comprise “novel by Mrs. Colleen” having a highest rank 522, “economic books” having a second highest rank 524, “fictional technology” having a third highest rank 526, and/or other candidate categories having respective rankings.

FIG. 6 illustrates an example 600 of a search engine website presenting a summary comprising a category 604. The search engine website may allow users to submit user search queries. In response to a user search query submission, the search engine website may return information relating to the user search query. In one example, a user may submit a user search query “Cash Management” 602 in an attempt to find additional information about a novel having a title “Cash Management”. The search engine website may present search results comprising a summary of a “Cash Management” that the user may have desired. The user may quickly reference the summary to determine whether the search engine website returned search results relating to the correct “Cash Management” the user desired. For example, the category 604 “novel by Mrs. Colleen” may be presented within the summary, and this may give an indication that the correct/relevant search results were returned. It may be appreciated that the category 604 may have been determined using at least some of one or more of the techniques described herein (e.g., method 100 of FIG. 1 and/or system 200 of FIG. 2).

FIG. 7 illustrates an example 700 of a search engine website providing a search query completion suggestion 704. The search engine website may allow users to submit user search queries. To aid the user in submitting accurate user search queries, the search engine website may provide the user with the search query completion suggestion 704. For example, user may input “Cash Management” 702 into the search engine website. To aid the user in narrowing down the search to what “Cash Management” the user is referencing (e.g., multiple novels may be named “Cash Management”, a movie may be named “Cash Management”, etc.), the search query completion suggestion 704 may be presented. The search query completion suggestion 704 may correspond to a category previously determined for the entity “Cash Management”. It may be appreciated that the category may have been determined using at least some of one or more of the techniques described herein (e.g., method 100 of FIG. 1 and/or system 200 of FIG. 2).

Still another embodiment involves a computer-readable medium comprising processor-executable instructions
configured to implement one or more of the techniques presented herein. An exemplary computer-readable medium that may be devised in these ways is illustrated in FIG. 8, wherein the implementation 800 comprises a computer-readable medium 816 (e.g., a CD-R, DVD-R, or a platter of a hard disk drive), on which is encoded computer-readable data 814. This computer-readable data 814 in turn comprises a set of computer instructions 812 configured to operate according to one or more of the principles set forth herein. In one such embodiment 800, the processor-executable computer instructions 812 may be configured to perform a method 810, such as the exemplary method 100 of FIG. 1, for example. In another such embodiment, the processor-executable instructions 812 may be configured to implement a system, such as the exemplary system 200 of FIG. 2, for example. Many such computer-readable media may be devised by those of ordinary skill in the art that are configured to operate in accordance with the techniques presented herein.

Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

As used in this application, the terms “component,” “module,” “system,” “interface,” and the like are generally intended to refer to a computer-related entity, either hardware, a combination of hardware and software, software, or software in execution. For example, a component may be, but is not limited to, a process running on a processor, a processor, an object, an executable, a thread of execution, a program, and/or a computer. By way of illustration, both an application running on a controller and the controller can be a component. One or more components may reside within a process and/or thread of execution and a component may be localized on one computer and/or distributed between two or more computers.

Furthermore, the claimed subject matter may be implemented as a method, apparatus, or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof to control a computer to implement the disclosed subject matter. The term “article of manufacture” as used herein is intended to encompass a computer program accessible from any computer-readable device, carrier, or media. Of course, those skilled in the art will recognize many modifications may be made to this configuration without departing from the scope or spirit of the claimed subject matter.

FIG. 9 and the following discussion provide a brief general description of a suitable computing environment to implement embodiments of one or more of the provisions set forth herein. The operating environment of FIG. 9 is only one example of a suitable operating environment and is not intended to suggest any limitation as to the scope of use or functionality of the operating environment. Example computing devices include, but are not limited to, personal computers, server computers, hand-held or laptop devices, mobile devices (such as mobile phones, Personal Digital Assistants (PDAs), media players, and the like), multiprocessor systems, consumer electronics, mini computers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

Although not required, embodiments are described in the general context of “computer readable instructions” being executed by one or more computing devices. Computer readable instructions may be distributed via computer readable media (discussed below). Computer readable instructions may be implemented as program modules, such as functions, objects, Application Programming Interfaces (APIs), data structures, and the like, that perform particular tasks or implement particular abstract data types. Typically, the functionality of the computer readable instructions may be combined or distributed as desired in various environments.

FIG. 9 illustrates an example of a system 910 comprising a computing device 912 configured to implement one or more embodiments provided herein. In one configuration, computing device 912 includes at least one processing unit 916 and memory 918. Depending on the exact configuration and type of computing device, memory 918 may be volatile (such as RAM, for example), non-volatile (such as ROM, flash memory, etc., for example) or some combination of the two. This configuration is illustrated in FIG. 9 by dashed line 914.

In other embodiments, device 912 may include additional features and/or functionality. For example, device 912 may also include additional storage (e.g., removable and/or non-removable) including, but not limited to, magnetic storage, optical storage, and the like. Such additional storage is illustrated in FIG. 9 by storage 920. In one embodiment, computer readable instructions to implement one or more embodiments provided herein may be in storage 920. Storage 920 may also store other computer readable instructions to implement an operating system, an application program, and the like. Computer readable instructions may be loaded in memory 918 for execution by processing unit 916, for example.

The term “computer readable media” as used herein includes computer storage media. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions or other data. Memory 918 and storage 920 are examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, Digital Versatile Disks (DVDs) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by device 912. Any such computer storage media may be part of device 912.

Device 912 may also include communication connection(s) 926 that allows device 912 to communicate with other devices. Communication connection(s) 926 may include, but is not limited to, a modem, a Network Interface Card (NIC), an integrated network interface, a radio frequency transmitter/receiver, an infrared port, a USB connection, or other interfaces for connecting computing device 912 to other computing devices. Communication connection(s) 926 may include a wired connection or a wireless connection. Communication connection(s) 926 may transmit and/or receive communication media.

The term “computer readable media” may include communication media. Communication media typically embodies computer readable instructions or other data in a “modulated data signal” such as a carrier wave or other trans-
port mechanism and includes any information delivery media. The term “modulated data signal” may include a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal.

Device 912 may include input device(s) 924 such as keyboard, mouse, pen, voice input device, touch input device, infrared cameras, video input devices, and/or any other input device. Output device(s) 922 such as one or more displays, speakers, printers, and/or any other output device may also be included in device 912. Input device(s) 924 and output device(s) 922 may be connected to device 912 via a wired connection, wireless connection, or any combination thereof. In one embodiment, an input device or an output device from another computing device may be used as input device(s) 924 or output device(s) 922 for computing device 912.

Components of computing device 912 may be connected by various interconnects, such as a bus. Such interconnects may include a Peripheral Component Interconnect (PCI), such as PCI Express, a Universal Serial Bus (USB), firewire (IEEE 1394), an optical bus structure, and the like. In another embodiment, components of computing device 912 may be interconnected by a network. For example, memory 918 may be comprised of multiple physical memory units located in different physical locations interconnected by a network.

Those skilled in the art will realize that storage devices utilized to store computer readable instructions may be distributed across a network. For example, a computing device 930 accessible via a network 928 may store computer readable instructions to implement one or more embodiments provided herein. Computing device 912 may access computing device 930 and download a part or all of the computer readable instructions for execution. Alternatively, computing device 912 may download pieces of the computer readable instructions, as needed, or some instructions may be executed at computing device 912 and some at computing device 930.

Various operations of embodiments are provided herein. In one embodiment, one or more of the operations described may constitute computer readable instructions stored on one or more computer readable media, which if executed by a computing device, will cause the computing device to perform the operations described. The order in which some or all of the operations are described should not be construed to imply that these operations are necessarily order dependent. Alternative ordering will be appreciated by one skilled in the art having the benefit of this description. Further, it will be understood that not all operations are necessarily present in each embodiment provided herein.

Moreover, the word “exemplary” is used herein to mean serving as an example, instance, or illustration. Any aspect or design described herein as “exemplary” is not necessarily to be construed as advantageous over other aspects or designs. Rather, use of the word exemplary is intended to present concepts in a concrete fashion. As used in this application, the term “or” is intended to mean an inclusive “or” rather than an exclusive “or”. That is, unless specified otherwise, or clear from context, “X employs A or B” is intended to mean any of the natural inclusive permutations. That is, if X employs A; X employs B; or X employs both A and B, then “X employs A or B” is satisfied under any of the foregoing instances. In addition, the articles “a” and “an” as used in this application and the appended claims may generally be construed to mean “one or more” unless specified otherwise or clear from context to be directed to a singular form.

Also, although the disclosure has been shown and described with respect to one or more implementations, equivalent alterations and modifications will occur to others skilled in the art upon a reading and understanding of this specification and the annexed drawings. The disclosure includes all such modifications and alterations and is limited only by the scope of the following claims. In particular regard to the various functions performed by the above described components (e.g., elements, resources, etc.), the terms used to describe such components are intended to correspond, unless otherwise indicated, to any component which performs the specified function of the described component (e.g., that is functionally equivalent), even though not structurally equivalent to the disclosed structure which performs the function in the herein illustrated exemplary implementations of the disclosure. In addition, while a particular feature of the disclosure may have been disclosed with respect to only one of several implementations, such feature may be combined with one or more other features of the other implementations as may be desired and advantageous for any given or particular application. Furthermore, to the extent that the terms “includes”, “having”, “has”, “with”, or variants thereof are used in either the detailed description or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising.”

What is claimed is:

1. A method for determining an entity category, comprising:
   extracting one or more potential categories and summary text relating to an entity from pre-labeled data;
   for respective potential categories, intersecting category words of a potential category with summary words within the summary text to determine a set of candidate categories;
   ranking the set of candidate categories based upon one or more ranking features; and
   determining an entity category from the ranked set of candidate categories having a desired rank.
2. The method of claim 1, comprising:
   presenting the entity category within a search result page in response to a user search query corresponding to the entity.
3. The method of claim 1, comprising:
   providing the entity category as a search query completion suggestion in response to user input comprising at least part of the entity’s name as a user search query.
4. The method of claim 1, the pre-labeled data corresponding to a web page.
5. The method of claim 4, the extracting comprising:
   extracting one or more potential categories from within a set of pre-labeled categories of the web page.
6. The method of claim 1, the extracting comprising:
   extracting a first occurring sentence within the pre-labeled data as the summary text.
7. The method of claim 1, the extracting comprising:
   extracting a first occurring paragraph within the pre-labeled data as the summary text.
8. The method of claim 1, the intersecting comprising:
   adding a potential category to the set of candidate categories based upon one or more category words within the potential category corresponding to one or more summary words within the summary text.
9. The method of claim 1, the intersecting comprising: performing a morphological analysis upon category words of a potential category to generate variation category words of the potential category; and intersecting the variation category words of the potential category with summary words within the summary text to determine the set of candidate categories;

10. The method of claim 1, the intersecting comprising: determining syntactical features of the summary text based upon at least one of a part-of-speech tagging, phrase chunking, and a natural-language processing technique; and excluding one or more potential categories from the set of candidate categories based upon the syntactical features.

11. The method of claim 1, the ranking comprising: assigning a desired rank to a candidate category based upon one or more candidate category words of the candidate category occurring after at least one of "is" and "was" within the summary text.

12. The method of claim 1, the ranking comprising: assigning an undesired rank to a candidate category based upon one or more category words of the candidate category occurring after at least one of "from" and "in" within the summary text.

13. The method of claim 1, the one or more ranking features corresponding to at least one of a position within summary text feature and a display size feature.

14. A system for determining an entity category, comprising:

an extraction component configured to:
extract one or more potential categories and summary text relating to an entity from pre-labeled data;

a filtering component configured to:
intersect category words of potential categories with summary words within the summary text to determine a set of candidate categories; and

a ranking component configured to:
rank the set of candidate categories based upon one or more ranking features; and
determine an entity category from the ranked set of candidate categories having a desired rank.

15. The system of claim 14, the filtering component configured to:
perform a morphological analysis upon category words of a potential category to generate variation category words of the potential category; and intersect the variation category words of the potential category with summary words within the summary text to determine the set of candidate categories.

16. The system of claim 14, the filtering component configured to:
determine syntactical feature of the summary text based upon at least one of a part-of-speech tagging, phrase chunking, a natural-language processing technique; and exclude one or more potential categories from the set of candidate categories based upon the syntactical features.

17. The system of claim 14, the ranking component configured to:
assign a desired rank to a candidate category based upon one or more category words of the candidate category occurring after at least one of "is" and "was" within the summary text.

18. The system of claim 14, the ranking component configured to:
assign an undesired rank to a candidate category based upon one or more category words of the candidate category occurring after at least one of "from" and "in" within the summary text.

19. The system of claim 14, comprising:
a presentation component configured to:
present the entity category in response to user input comprising at least part of the entity's name.

20. A method for determining an entity category comprising:
extracting one or more potential categories from an article of a web page;
extracting summary text from at least a portion of a first paragraph of the article;
for respective potential categories, intersecting category words of a potential category with summary words within the summary text to determine a set of candidate categories;
excluding potential categories from the set of candidate categories based upon syntactical features of the summary text;
ranking the set of candidate categories based upon at least one of a position within summary text feature and a display size feature; and
determining an entity category from the ranked set of candidate categories having a desired rank.

* * * * *